Insect Molecular Biology (2005) **14**(2), 121–136

Midgut and salivary gland transcriptomes of the arbovirus vector *Culicoides sonorensis* (Diptera: Ceratopogonidae)

C. L. Campbell*, K. A. Vandyke†, G. J. Letchworth*, B. S. Drolet*, T. Hanekamp‡ and W. C. Wilson*

**USDA, ARS, Arthropod-Borne Animal Diseases Research Laboratory, College of Agriculture, Department 3354, 1000 E. University, Laramie, WY, USA; and †Department of Zoology and Physiology, and ‡Department of Molecular Biology, University of Wyoming, Laramie, WY, USA*

Abstract

Numerous *Culicoides* spp. are important vectors of livestock or human disease pathogens. Transcriptome information from midguts and salivary glands of adult female *Culicoides sonorensis* provides new insight into vector biology. Of 1719 expressed sequence tags (ESTs) from adult serum-fed female midguts harvested within 5 h of feeding, twenty-eight clusters of serine proteases were derived. Four clusters encode putative iron binding proteins (FER1, FERL, PXDL1, PXDL2), and two clusters encode metalloendopeptidases (MDP6C, MDP6D) that probably function in bloodmeal catabolism. In addition, a diverse variety of housekeeping cDNAs were identified. Selected midgut protease transcripts were analysed by quantitative real-time PCR (q-PCR): TRY1_115 and MDP6C mRNAs were induced in adult female midguts upon feeding, whereas TRY1_156 and CHYM1 were abundant in midguts both before and immediately after feeding. Of 708 salivary gland ESTs analysed, clusters representing two new classes of protein families were identified: a new class of D7 proteins and a new class of Kunitz-type protease inhibitors. Additional cDNAs representing putative immunomodulatory proteins were also identified: 5′ nucleotidases, antigen 5-related proteins, a hyaluronidase, a platelet-activating factor acetylhydrolase, mucins and several immune response cDNAs. Analysis by q-PCR showed that all D7 and Kunitz domain transcripts tested were highly enriched in female heads compared with other tissues and were generally absent from males. The mRNAs of two additional protease inhibitors, TFPI1 and TFPI2, were detected in salivary glands of paraffin-embedded females by *in situ* hybridization.

Keywords: expressed sequence tag, EST, haematophagous, gut, vector biology.

Introduction

Of about 1800 *Culicoides* species world-wide, a small subset are economically important vectors of livestock arthropod-borne viruses (arboviruses) or medically important vectors of filarial or viral pathogens (reviewed in Mellor *et al*., 2000). The midgut and salivary gland tissues are of keen interest to vector biologists for study of the propagation, dissemination and transmission of arboviruses by these biting midges. According to dogma, the midgut milieu either restricts or actuates arbovirus proliferation in the insect vector. Subsequent virus dissemination to the salivary glands is a prerequisite for bite transmission to a naïve mammalian host.

As pool feeders, *Culicoides* secrete pharmacologically important compounds into the bite site during blood-feeding to prevent coagulation and encourage vasodilation (Perez de Leon & Tabachnick, 1996; Mckeever *et al*., 1997; Perez de Leon *et al*., 1997, 1998). These factors may also alter the host's innate and acquired immune responses (Perez de Leon & Tabachnick, 1996; Perez de Leon *et al*., 1997, 1998; Limesand *et al*., 2000, 2003). Furthermore, saliva pharmacological factors of other disease vectors potentiate infection of a wide variety of pathogens, including arboviruses (Jones *et al*., 1992), protozoa (Titus & Ribeiro, 1988; Norsworthy *et al*., 2004) and bacteria (Gillespie *et al*., 2001).

The growing dataset of genomic information and web-based toolkits, such as FlyBase, for well-described dipterans (Adams *et al*., 2000; Myers *et al*., 2000; Holt *et al*., 2002; Zdobnov *et al*., 2002; Ribeiro, 2003) has greatly facilitated study of lesser known vector insects, such as *Culicoides sonorensis*. Furthermore, the growing transcriptome datasets from specific tissues of blood-sucking dipterans such as mosquitoes, sand flies, and tsetse flies (Charlab *et al*., 1999; Lehane *et al*., 2003; Valenzuela *et al*., 2003; Calvo *et al*., 2004) allow more accurate sequence annotation for

doi: 10.1111/j.1365-2583.2004.00537.x

Received 15 June 2004; accepted after revision 1 October 2004. Correspondence: Dr Corey L. Campbell, USDA, ARS, ABADRL, College of Agriculture, Department 3354, 1000 E. University Ave., Laramie, WY 82071, USA. Tel.: +1 307 766 3626; fax: +1 307 766 3500; e-mail: camcorey@uwyo.edu

other families of disease vectors. In turn, analysis of the *C. sonorensis* transcriptome should provide insight into the convergent evolution of haematophagous insects and facilitate functional genomics studies of vector–pathogen–host relationships.

Until now, very little genomic information has been available for *Culicoides* biting midges. Here, we report analysis of expressed sequence tags (ESTs) from salivary glands and midguts of *C. sonorensis*. These cDNA datasets reveal the complexity of blood-feeding and digestive processes in this vector insect.

Results and discussion

Midgut and salivary gland transcriptomes

In this study, 1719 ESTs from a serum-fed adult female midge midgut cDNA library and 708 ESTs from an adult female salivary gland cDNA library were analysed. For putative functional assignments, ESTs were compared with those in the public domain by sequence similarity search, tblastx or blastx (Altschul *et al.*, 1990). Midgut ESTs were grouped by Gene Ontology biological process (Fig. 1A) (Ashburner *et al.*, 2000), and salivary gland ESTs were categorized by functional group (Fig. 1B). Because few public sequences were available for the family Ceratopogonidae, nucleotide comparisons were largely uninformative. This resulted in translated sequence similarity scores and corresponding E values for many *Culicoides* cDNAs to be lower than those reported for other dipteran families.

The midgut and salivary gland EST collections correspond to two unique metabolic milieus. The midgut EST collection was prepared from poly A^+ RNA of female midges 2–5 h following a serum meal. Therefore, most ESTs represent digestive enzymes and conserved metabolic factors. The adult female salivary gland cDNA library was prepared from midges that had been provided with sugar water as a food source. Library construction required the use of total RNA and a PCR-based cloning strategy, and an initial assessment of the EST collection revealed a large proportion of novel genes. In order to prevent reporting of sequences resulting from PCR errors or chimeras generated during cDNA library construction, novel EST singletons with no identifiable Pfam protein domains (Bateman *et al.*, 2004) and no apparent open reading frame (ORF) were excluded from further analysis. Therefore, of 1259 salivary gland ESTs generated, 708 are reported here. Even so, relaxation of analytical parameters was required to categorize many salivary gene clusters, as salivary factors in *Culicoides* may have evolved independently of those of other haematophagous arthropods. Because of the inherent differences between these two EST collections and the tissues they represent, each dataset is presented in such a way as to reflect its unique characteristics.

Midgut cDNAs of interest

As indicated in Fig. 1A, a variety of diverse biological processes are represented in the midgut EST collection; however, in this report we have focused on selected functional groups. Of 1719 serum-fed adult female midgut ESTs analysed, the most abundant functional group contains 600 ESTs and encodes for putative proteins of proteolytic and peptidolytic functions. Of these, 371 encode about twenty-eight clusters of serine proteases, mostly trypsin or chymotrypsin. It is not yet clear whether any of these proteolytic enzymes might interact with ingested viruses, as has been reported for other arthropods (Nakazawa *et al.*, 2004).

The 197 clusters in the protein metabolism category of Fig. 1A account for about 11% of the total number of midgut ESTs analysed. Components of the protein translation machinery are represented by seventy-eight EST clusters in this category. These include major translation elongation factors and a variety of ribosomal protein subunits, as well as tRNA synthetases and transaminases. In contrast, the salivary gland EST collection contains just forty-seven clusters in the protein metabolism category, about 6% of the 708 sequences reported here. Of these, twenty-two ribosomal protein subunits are represented (data not shown).

Table 1 depicts EST clusters for selected functional groups: defence, mitochondrial function, heme catabolism and major digestive proteases. The EST clusters representing putative immune response genes are expected to act in antibacterial immunity. A group of thirty-nine ESTs, representing twenty-six clusters, code for proteins localized to the mitochondrial compartment (Adams *et al.*, 2000; Ashburner *et al.*, 2000). These include a full range of components associated with mitochondrial function, including respiratory chain complex components and transporters, such as those that shuttle ATP/ADP, phosphate or citrate across the mitochondrial membrane. In contrast, just eleven clusters of the salivary cDNA collection code for proteins localized to mitochondria. Most were constituents of the hydrogen-transporting ATPase complex (data not shown), perhaps contributing to cell membrane potentials required for high levels of secretory activity (Berridge *et al.*, 1984; Zimmermann, 2000).

Seven midgut EST clusters were identified that may provide protection against the toxic effects of heme during bloodmeal digestion. These include two clusters of metal-dependent proteases (MDP6C, MDP6D), as well as ferrous or heme binding proteins: FER1, FERL, PXDL1 and PXDL2 (Table 1). In addition, cytochrome b5 in the mitochondrial category may also participate in heme catabolism (Wang *et al.*, 2003). Interestingly, an additional cluster, THIL, was also identified that shows weak sequence similarity to thrombin inhibitors. It may help to prevent coagulation of the bloodmeal within the midgut during blood catabolism.

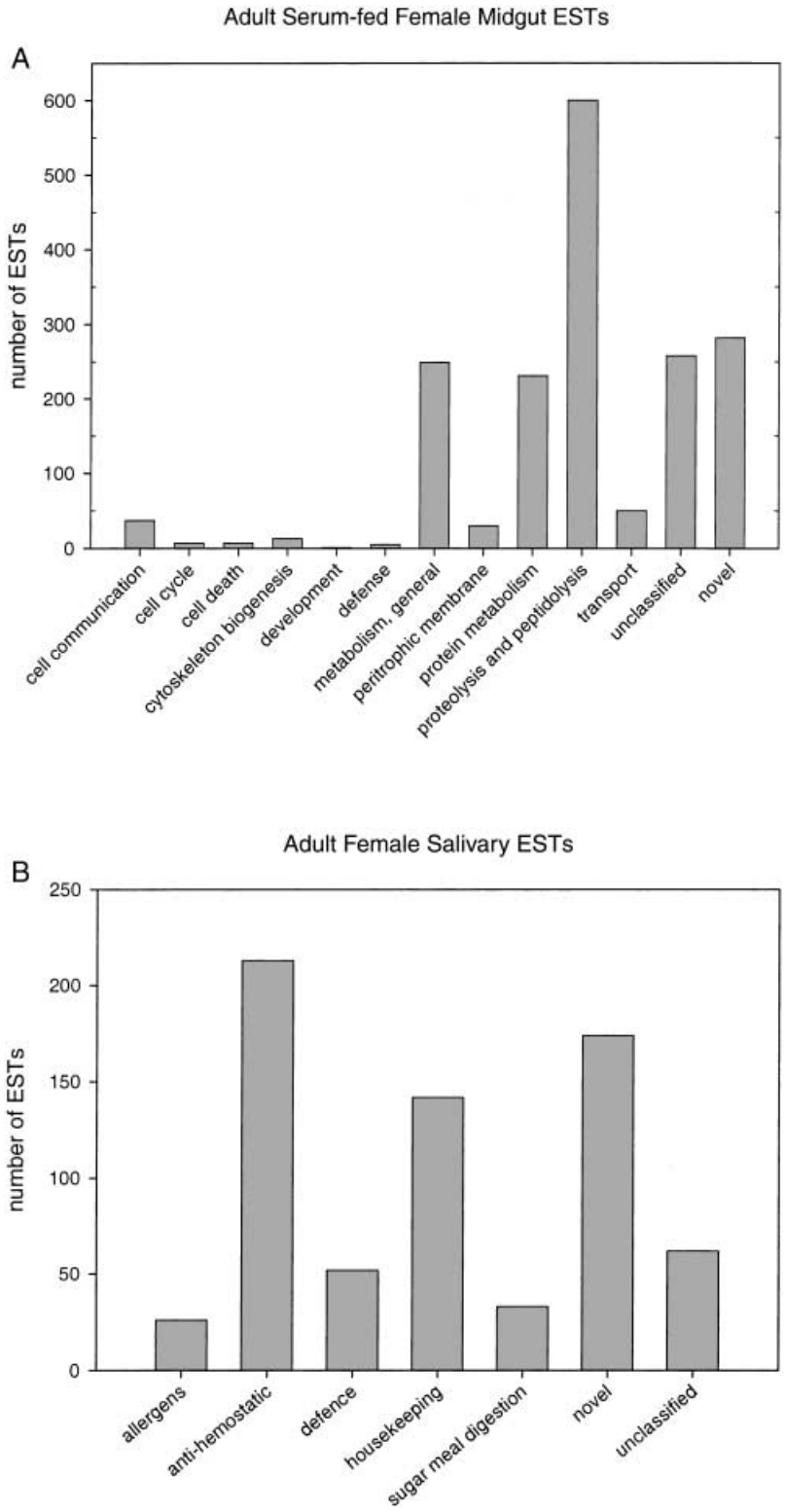


Figure 1. Expressed sequence tag clusters organized by functional group. ESTs were analysed for sequence similarity of the conceptually translated protein to known proteins by blastx. Unclassified ESTs are similar to public domain sequences of undetermined function. (A) *Culicoides* serum-fed adult female midgut ESTs (n = 1719) categorized by Gene Ontology biological process™. Novel ESTs are those that were not similar (E value < 1E-06) to GenBank sequences. (B) Adult female salivary gland ESTs (n = 708) categorized by functional group.

Salivary gland cDNAs of interest

A portion of the salivary gland EST clusters depicted in Fig. 1B are listed by functional category in Table 2. Functional groups include putative allergens and other immunomodulatory factors, a novel class of Kunitz-type protease inhibitors and a new class of D7 proteins. As indicated in the table, many of these cDNAs have signal peptides,

Table 1. Selected *Culicoides sonorensis* midgut EST clusters

No. of sequences	Cluster (accession no.)	Best match	E value	Putative function
Defence				
1	CsmgEST00249 (CK149743)	17864128	4.00E-71	Saposin-related
1	CsmgEST00407 (CK149840)	18568294	3.00E-34	Gram negative bacteria binding protein
1	CsmgEST01070 (CK150177)	31238563	6.00E-15	Chitinase-like
1	CsmgEST01076 (CK150180)	5476007	2.00E-10	Toll-like
1	CsmgEST01523 (CK150441)	30176016	5.00E-63	Peptidoglycan recognition protein
1	CsmgEST02328 (CK150900)	25013110	1.00E-25	Fatty acid carrier
Mitochondrial function				
2	MPCP (CK150949, CK150912)	31208744	1.00E-99	Phosphate transporter
2	HSC70A (CK150030, CK150370)	31241094	0.00E+00	Chaperone
2	HSC70B (CK150335, CK150967)	28557576	0.00E+00	Chaperone
1	CsmgEST01877 (CK150647)	231710	2.00E-45	Citrate transporter
3	CYTB51 (AY752801)	28573284	2.10E-14	Cytochrome b5
1	CsmgEST00326 (CK149792)	28416333	4.00E-57	Cytochrome *c* oxidase polypeptide VbI
1	CsmgEST00175 (CK149695)	246433	3.00E-28	Cytochrome *c* oxidase
1	CsmgEST00297 (CK149772)	24658560	4.00E-47	Hydrogen-exporting ATPase
1	CsmgEST02347 (CK150913)	2465160	2.00E-37	Hydrogen-exporting ATPase
1	CsmgEST01492 (CK150422)	2476268	2.00E-47	Hydrogen-exporting ATPase
4	ANT (AY752790)	28315874	1.50E-32	Adenine nucleotide translocase
1	CsmgEST01876 (CK150646)	17647687	1.00E-37	mRPS26: mitochondrial ribosomal protein S26
2	NDH4 (CK150982, CK149973)	6103662	3.40E-57	NADH dehydrogenase (ubiquinone)
2	PRX1 (CK150109, CK150090)	31203510	2.00E-88	Peroxidase
1	CsmgEST01773 (CK150588)	20792390	2.00E-69	Thioredoxin reductase
2	TCC1 (CK150128)	25012932	2.0 E-19	Tricarboxylate carrier
2	PORIN1 (CK150095, CK150655)	22004649	0.00E+00	Voltage-dependent Ion-selective channel activity
1	CsmgEST02027 (CK150730)	31216089	3.00E-041	Carnitine/acylcarnitine translocase
1	CsmgEST00998 (CK150136)	24657533	3.00E-55	Transporter
1	CsmgEST01192 (CK150256)	5834918	2.00E-14	NADH dehydrogenase subunit 3
1	CsmgEST0450 (CK149864)	28269709	1.00E-55	ADP/ATP translocase
1	CsmgEST01681 (CK150541)	2275935	1.00E-87	NADH dehydrogenase
1	CsmgEST01677 (CK150538)	7290739	4.00E-22	NADH dehydrogenase
2	Csmg00391 (CK151005, CK151066)	21064230	9.00E-82	NADH dehydrogenase subunit 5
1	CsmgEST02223 (CK150843)	31210517	1.00E-45	NADH-ubiquinone oxidoreductase
1	CsmgEST00701 (CK149990)	30175068	5.00E-85	Succinate-Coenzyme A ligase, ADP-forming, beta
Heme catabolism				
2	FERL (CK150345, CK150348)	16902029	3.00E-16	Ferrous iron binding
4	FER1*† (AY752804)	37727747	9.00E-030	Ferritin light chain, ferrous iron binding
18	MDP6C*† (AY752808)	31231535	5.00E-63	Metalloendopeptidase
7	MDP6D† (AY752809)	31231535	3.00E-53	Metalloendopeptidase
2	PXDL1 (CK150144)	31213520	6.00E-11	Heme or iron ion binding; peroxidase
2	PXDL2 (CK151107, CK150010)	31213520	1.00E-11	Heme or iron ion binding; peroxidase
7	THIL*† (AY752845)	31225481	3.00E-10	Thrombin inhibitor-like
Major digestive proteases				
46	CHYM1*† (AY752796)	31240394	4.00E-52	Chymotrypsin
25	CPA1A*† (AY752797)	31221729	0	Metallocarboxypeptidase
25	CPA1B*† (AY752798)	31221729	0	Metallocarboxypeptidase
31	CPA3*† (AY752799)	28574957	0	Metallocarboxypeptidase
7	TRY1_049*† (AY752846)	479774	2.00E-060	Serine protease
10	TRY1_102*† (AY752847)	31247956	3.00E-042	Serine protease
25	TRY1_115*† (AY752848)	31222378	6.00E-049	Serine protease
54	TRY1_156*† (AY752857)	479774	8.00E-073	Serine protease

'Cluster' indicates consensus cDNA or EST with associated GenBank accession number(s). 'Novel' indicates that sequence similarity to known proteins had an E value > 1E-06. 'Best match' indicates the gi number of the most similar annotated cDNA in GenBank.
*Complete open reading frame.
†Signal peptide (SignalP; Nielsen *et al.*, 1997). Signal P cutoff 0.5.

suggesting they are secreted proteins. Commonly, 'novel' EST clusters are those that have a sequence similarity E value of greater than 10E-06 to public domain sequences. However, because the independent evolution of *Culicoides* salivary factors has resulted in genes with weak similarity to those of other haematophagous insects, some entries have similarity E values as weak as 0.001. In these cases, the presence of a predicted protein domain provides preliminary indication of proposed function. Functional studies will be required to confirm these assignments.

Table 2. Selected salivary EST clusters

No. of sequences	Cluster (accession no.)	Best match	E value	Putative function	Domain or feature	E value
Allergens						
3	HGS1 (AY603562)	1346322	7.00E-23	Hyaluronidase	Glyco_hydro_56	6.00E-37
1	CssgEST00738 (CN612973)	7300721	1.00E-07	Hyaluronidase-like		
20	AG5A*† (AY603555)	18568308	5.00E-45	Antigen 5-related	SCP	7.10E-35
2	AG5B*† (AY603556)	24641974	2.00E-35	Antigen 5-related	SCP	1.70E-17
Platelet aggregation inhibitors						
1	CssgEST00590* (CN612855)	19572986	1.00E-36	5′ nucleotidase	Metallophos	2.5E-05
1	CssgEST01324*† (CN613420)	19572985	5.00E-36	5′ nucleotidase	Metallophos	4.2E-08
1	CssgEST00543 (CN612816)	31200353	1.00E-67	Platelet-activating factor acetylhydrolase		
2	COP1 (AY603641)	5733713	8.00E-07	5′ nucleotidase-like	5_nucleotid_C	9.7E-09
D7 family						
30	Cssg00004*† (AY603569)	Novel			PBP_GOBP	0.0021
13	Cssg00015*† (AY603571)	16225992	2.50E-01	D7Bclu1	PBP_GOBP	3.70E-05
5	Cssg00211*† (AY603587)	15963511	0.37	Protein precursor	PBP_GOBP	0.00021
2	Cssg00331*† (AY603595)	24651102	0.026	OBP-like	PBP_GOBP	8.00E-05
21	Cssg00642A*† (AY603607)	19922636	8.30E-01	OBP-like	PBP_GOBP	0.00021
12	Cssg00642B*† (AY603608)	Novel			PBP_GOBP	0.00013
6	Cssg00666A*† (AY603626)	31747535	0.12	PBP-like	PBP_GOBP	4.20E-06
4	Cssg00666B*† (AY603627)	10129673	0.05	PBP-like	PBP_GOBP	6.50E-06
2	Cssg00673A (AY603633)	27414107	0.33	OBP-like	PBP_GOBP	2.20E-07
3	Cssg00673B*† (AY603634)	38350687	0.59	Cu pip. D7 short-like	PBP_GOBP	2.30E-07
1	CssgEST00850 (CN613065)	17981809	0.12	OBP-like	PBP_GOBP	0.083
1	CssgEST00496 (CN612779)	27679000	5.30E-01	Olfactory receptor-like		
Other coagulation inhibitors						
10	TFPI1*† (AY603642)	8394443	1.00E-10	Kunitz protease inhibitor-like	Kunitz_BPTI	0.00016
12	TFPI2*† (AY603643)	2148085	6.00E-24	Kunitz protease inhibitor	Kunitz_BPTI	6.80E-20
15	TFPI3 (AY603644)	Novel		Kunitz protease inhibitor	Kunitz_BPTI	6.10E-14
12	Cssg00129*† (AY603581)	22901764	4.00E-04	Kunitz protease inhibitor-like	Kunitz_BPTI	0.0015
19	Cssg00654A (AY603615)	17558476	2.00E-03	Thrombospondin-like	Kunitz_BPTI	0.021
5	Cssg00654B (AY603616)	17558476	2.00E-03	Thrombospondin-like	Kunitz_BPTI	0.021
15	Cssg00660A*† (AY603620)	22901764	0.004	Protease inhibitor-like	Venom trypsin inhibitor	0.01
3	Cssg00660B (AY603621)	22901764	n/a	Protease inhibitor-like	Venom trypsin inhibitor	0.01
1	CssgEST01066 (CN613232)	24645189	1.00E-09	Serine protease inhibitor	Kazal	0.0042
Defence						
1	CssgEST01396 (CN613470)	7511780	2.00E-22	Saposin-related	SapB_2; SapB_1	0.00015; 0.00028
1	CssgEST00553*† (CN612823)	24661689	9.00E-59	Chitin-binding	CBM_14	1.20E-12
3	Cssg00285* (AY603592)	Novel		Saposin-related	SapA	0.22
7	Cssg00679A (AY603639)	24580947	7.00E-16	Unclassified	Penaeidin	0.074
4	Cssg00679B (AY603640)	24580947	1.00E-16	Unclassified	Penaeidin	0.058
Sugar meal digestion						
10	AMY1*† (AY603557)	7435318	4.00E-62	Amylase	Alpha amylase_C	1.40E-32
22	MAL1*† (AY603565)	17944413	0.00E+00	Maltase	Alpha-amylase	5.30E-107
1	CssgEST01437 (CN613505)	31206819	7.00E-58	Aldolase		
Proteases						
7	LTRYP3A*† (AY603563)	31236527	2.00E-54	Serine protease	Trypsin	5.70E-62
9	LTRYP3B*† (AY603564)	28566192	2.00E-40	Serine protease	Trypsin	3.70E-61
4	Cssg00128 (AY603580)	Novel		Possible membrane carboxypeptidase	Membrane carboxypeptidase	0.007 (CDD)
Serine/rhreonine-rich low complexity proteins						
6	Cssg00019*† (AY603573)	Novel		Mucin	O-glycosylated (NetoGlyc)	
7	Cssg00021*† (AY603574)	Novel		Mucin	O-glycosylated (NetoGlyc)	
9	Cssg00656A*† (AY603617)	Novel		Mucin	O-glycosylated (NetoGlyc)	
12	Cssg00656B*† (AY603618)	Novel		Mucin	O-glycosylated (NetoGlyc)	
2	Cssg00656C*† (AY603619)	Novel		Mucin	O-glycosylated (NetoGlyc)	
Novel or unclassified secreted proteins						
10	Cssg00016*† (AY603572)	Novel			Resistin	0.18
5	Cssg00082*† (AY603579)	Novel				
7	Cssg00179*† (AY603583)	Novel				
11	Cssg00428*† (AY603597)	Novel			7tm_5	0.05
2	Cssg00436† (AY603599)	Novel				
9	Cssg00644A*† (AY603609)	Novel				

Table 2. (*Continued*)

No. of sequences	Cluster (accession no.)	Best match	E value	Putative function	Domain or feature	E value
3	Cssg00644B*† (AY603610)	Novel				
6	Cssg00651A*† (AY603613)	Novel			DUF388	0.16
5	Cssg00651B*† (AY603614)	Novel				
6	Cssg00668A*† (AY603628)	Novel			L71	0.017
4	Cssg00668B*† (AY603629)	Novel			L71	0.0004
2	Cssg00668C*† (AY603630)	Novel				
7	Cssg00675A*† (AY603635)	Novel				
11	Cssg00675B*† (AY603636)	Novel				
2	Cssg00433*† (AY603598)	31242735	2.00E-18	Unclassified		
1	CssgEST00135*† (CN613684)	18568316	9.00E-34	Unclassified	SCP	4.00E-10
1	CssgEST00705*† (CN613707)	18568316	4.00E-32	Unclassified	SCP	1.70E-08
1	CssgEST00834 (CN613051)	18568316	2.00E-35	Unclassified	SCP	1.20E-15
1	CssgEST00925 (CN613128)	18568308	1.00E-32	Unclassified	SCP	5.50E-33
1	CssgEST01560 (CN613736)	18568316	1.00E-34	Unclassified	SCP	2.60E-15
Other novel or unclassified cDNAs						
28	Cssg00011* (AY603570)	Novel			Laminin II	0.029
7	Cssg00162*† (AY603582)	Novel			Diptheria_T	0.25
3	Cssg00210* (AY603586)	Novel			Tuberin	0.4
17	Cssg00637A* (AY603602)	Novel			NfeD	0.088
13	Cssg00637B* (AY603603)	Novel			NfeD	0.097
3	Cssg00664A* (AY603624)	Novel			Kunitz_BPTI	0.34
2	Cssg00664B* (AY603625)	Novel			Kunitz_BPTI	0.13
10	Cssg00671A (AY603631)	Novel				
5	Cssg00671B (AY603632)	Novel				
18	Cssg00677A (AY603637)	Novel				
10	Cssg00677B (AY603638)	Novel				
2	Cssg00025 (AY603575)	24580947	6.00E-06	Unclassified		
3	Cssg00181 (AY603584)	18858077	1E-10	Unclassified		
3	Cssg00254* (AY603589)	24580947	7.00E-06	Unclassified		
2	Cssg00302 (AY603594)	18858077	7.00E-22	Unclassified		
8	Cssg00374 (AY603596)	24580947	8.00E-07	Unclassified		
1	Cssg00437* (CN612744)	31239005	2.00E-06	Unclassified		
2	Cssg00031 (AY603576)	27363942	3.00E-12	Unclassified		

Listed with the most similar known protein by blastx search, corresponding E value and presence of any putative signal peptide (SignalP website) (Nielsen *et al*., 1997), domain (Pfam website) (Bateman *et al*., 2004) or other feature. Except where indicated, all domains were determined using Pfam. 'Cluster' indicates consensus cDNA name with associated GenBank accession number. 'Best match' indicates the gi number of the most similar annotated protein in GenBank nr database, blastx search. 'Novel' indicates that sequence similarity to known proteins had an E value > 0.001. Pfam protein domain E value cutoff 0.5. CDD, Conserved Domain Database (Marchler-Bauer *et al*., 2003); NetoGlyc (Hansen *et al*., 1998).

*Complete open reading frame.

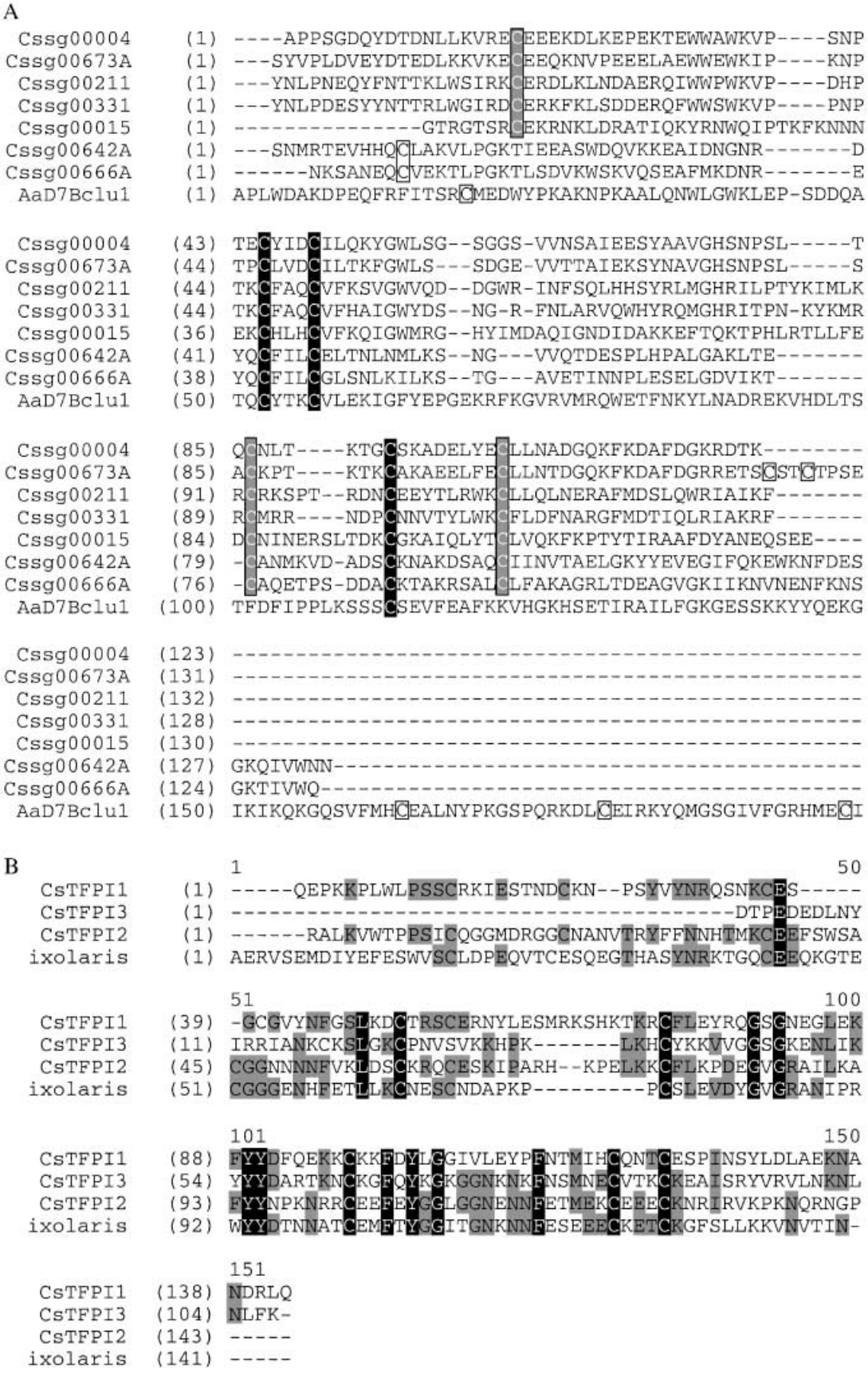
†Signal peptide (SignalP; Nielsen *et al*., 1997).

List of protein domain abbreviations: CBM_4, chitin binding domain; DUF, Domain unknown function; Kazal, Kazal-type serine protease inhibitor; Kunitz_BPTI, Kunitz/Bovine pancreatic trypsin inhibitor; 5_nucleotid_C, 5′ nucleotidase; 7tm_5, seven transmembrane; Glyco_hydro_56, glycoside hydrolase; L71, L71 family; metallophos, metallophosphoesterase; NfeD, Nodulation efficiency protein D; PBP_GOBP, Pheromone binding protein – general odorant binding proteins; Penaeidin, antimicrobial peptide; SapB_1, SapB_2, SapA, saposin type domains; SCP, sperm coating protein.

Allergens

Culicoides allergens cause type I hypersensitivity in horses, known as sweet itch (Baker & Quinn, 1978; Quinn *et al*., 1983), and may cause similar responses in other animals (Yeruham *et al*., 1993). The two major clusters in Table 2 represent putative allergens similar to those of Hymenoptera and Nematocera. The predicted translation of the first cluster, HGS, encodes a hyaluronidase containing a glycoside hydrolase 56 protein domain, as does the most similar orthologue, a venom allergen hyaluronidase from a hornet (Lu *et al*., 1995). Hyaluronidase enzyme activity is common to pool-feeding arthropods, such as leeches, sandflies and black flies and may facilitate enlargement of the lesion during feeding (Budds *et al*., 1987; Charlab *et al*., 1999; Ribeiro *et al*., 2000a; Cerna *et al*., 2002). Small molecular weight hyaluronidases induce macrophage activity during the inflammatory response (Mckee *et al*., 1996, 1997). Ribeiro *et al*. (2000a) suggested that this could promote establishment of an arbovirus infection in the mammalian host if the

Figure 2. (A) Mature *Culicoides* D7-related proteins share conserved cysteines with D7Bclu1 from *Aedes aegypti*. All cysteines are boxed. Identical cysteines are shaded in black; *Culicoides* conserved cysteines are in grey. Alignments were made using the VNTI multiple alignment tool, gonnet matrix, gap opening penalty 10, gap extension penalty 0.2. Other alleles not shown, Cssg00673B, Cssg00642B, and Cssg00666B, contain amino acid substitutions that do not alter the spacing of cysteines. (B) Alignment of mature TFPI proteins and the tick orthologue, ixolaris, gonnet matrix, gap opening penalty 10, gap extension penalty 0.05. Identical amino acis are in shaded in black; conserved amino acids are in grey.


A
Cssg00004 (1) ----APPSGDQYDTDNLLKVREEEKDLKEPEKTEWWAWKVP----SNP
Cssg00673A (1) ---SYVPLDVEYDTEDLKKVKEEEQKNVPEEELAEWWEWKIP----KNP
Cssg00211 (1) ---YNLPNEQYFNTTKLWSIRKERDLKLNDAERQIWWPWKVP----DHP
Cssg00331 (1) ---YNLPDESYYNTTRLWGIRDERKFKLSDDERQFWWSWKVP----PNP
Cssg00015 (1) ----------GTRGTSREKRNKLDRATIQKYRNWQIPTKFKNNN
Cssg00642A (1) ---SNMRTEVHHQCLAKVLPGKTIEEASWDQVKKEAIDNGNR-------D
Cssg00666A (1) ------NKSANEQCVEKTLPGKTLSDVKWSKVQSEAFMKDNR-------E
AaD7Bclu1 (1) APLWDAKDPEQFRFITSRCMEDWYPKAKNPKAALQNWLGWKLEP-SDDQA
Cssg00004 (43) TECYIDCILQKYGWLSG--SGGS-VVNSAIEESYAAVGHSNPSL-----T
Cssg00673A (44) TPCLVDCILTKFGWLS---SDGE-VVTTAIEKSYNAVGHSNPSL-----S
Cssg00211 (44) TKCFAQCVFKSVGWVQD--DGWR-INFSQLHHSYRLMGHRILPTYKIMLK
Cssg00331 (44) TKCFAQCVFHAIGWYDS--NG-R-FNLARVQWHYRQMGHRITPN-KYKMR
Cssg00015 (36) EKCHLHCVFKQIGWMRG--HYIMDAQIGNDIDAKKEFTQKTPHLRTLLFE
Cssg00642A (41) YQCFILCELTNLNMLKS--NG---VVQTDESPLHPALGAKLTE-------
Cssg00666A (38) YQCFILCGLSNLKILKS--TG---AVETINNPLESELGDVIKT-------
AaD7Bclu1 (50) TQCYTKCVLEKIGFYEPGEKRFKGVRVMRQWETFNKYLNADREKVHDLTS
Cssg00004 (85) QCNLT----KTGCSKADELYECLLNADGQKFKDAFDGKRDTK--------
Cssg00673A (85) ACKPT----KTKCAKAEELFECLLNTDGQKFKDAFDGRRETSCSTCTPSE
Cssg00211 (91) RCRKSPT--RDNCEEYTLRWKCLLQLNERAFMDSLQWRIAIKF-------
Cssg00331 (89) RCMRR----NDPCNNVTYLWKCFLDFNARGFMDTIQLRIAKRF-------
Cssg00015 (84) DCNINERSLTDKCGKAIQLYTCLVQKFKPTYTIRAAFDYANEQSEE----
Cssg00642A (79) -CANMKVD-ADSCKNAKDSAQCIINVTAELGKYYEVEGIFQKEWKNFDES
Cssg00666A (76) -CAQETPS-DDACKTAKRSALCLFAKAGRLTDEAGVGKIIKNVNENFKNS
AaD7Bclu1 (100) TFDFIPPLKSSSCSEVFEAFKKVHGKHSETIRAILFGKGESSKKYYQEKG
Cssg00004 (123) ----------
Cssg00673A (131) ----------
Cssg00211 (132) ----------
Cssg00331 (128) ----------
Cssg00015 (130) ----------
Cssg00642A (127) GKQIVWNN----------
Cssg00666A (124) GKTIVWQ----------
AaD7Bclu1 (150) IKIKQKGQSVFMHCEALNYPKGSPQRKDLCEIRKYQMGSGIVFGRHMECI
B
1 50
CsTFPI1 (1) -----QEPKKPLWLPSSCRKIESTNDCKN--PSYVYNRQSNKCES-----
CsTFPI3 (1) ----------DTPEDEDLNY
CsTFPI2 (1) ------RALKVWTPPSICQGGMDRGGCNANVTRYFFNNHTMKCEEFSWSA
ixolaris (1) AERVSEMDIYEFESWVSCLDPEQVTCESQEGTHASYNRKTGQCEEQKGTE
51 100
CsTFPI1 (39) -GCGVYNFGSLKDCTRSCERNYLESMRKSHKTKRCFLEYRQGSGNEGLEK
CsTFPI3 (11) IRRIANKCKSLGKCPNVSVKKHPK-------LKHCYKKVVGGSGKENLIK
CsTFPI2 (45) CGGNNNNFVKLDSCKRQCESKIPARH--KPELKKCFLKPDECVGRAILKA
ixolaris (51) CGGGENHFETLLKCNESCNDAPKP---------PCSLEVDYGVGRANIPR
101 150
CsTFPI1 (88) FYYDFQEKKCKKFDYLGGIVLEYPFNTMIHCQNTCESPINSYLDLAEKNA
CsTFPI3 (54) YYYDARTKNCKGFQYKGKGGNKNKFNSMNECVTKCKEAISRYVRVLNKNL
CsTFPI2 (93) FYYNPKNRRCEEFEYGGLGGNENNFETMEKCEEECKNRIRVKPKNQRNGP
ixolaris (92) WYYDTNNATCEMFTYGGITGNKNNFESEEECKETCKGFSLLKKVNVTIN-
151
CsTFPI1 (138) NDRLQ
CsTFPI3 (104) NLFK-
CsTFPI2 (143) -----
ixolaris (141) -----

pathogen in question infects macrophages, thus aiding dissemination.

A second major salivary EST cluster in this category encodes a predicted protein similar to antigen 5-related protein of *Aedes aegypti*. As with hyaluronidases, antigen 5 proteins are common to venom allergens of Hymenoptera and salivary factors of Nematocera. In addition, antigen 5 proteins have been proposed to belong to the pathogenesis-related protein superfamily (Henriksen *et al*., 2001). Although they have been identified in the salivary glands of several disease vector species, any clues to function remain unknown (Li *et al*., 2001; Francischetti *et al*., 2002b; Valenzuela *et al*., 2002c; Ribeiro *et al*., 2004).

Platelet aggregation inhibitors

Apyrases are found in a variety of haematophagous insects of the suborder Nematocera (Champagne *et al*., 1995; Charlab *et al*., 1999); they facilitate blood-feeding through the inhibition of ADP-mediated platelet aggregation. Divalent cation-dependent apyrase activity was previously found in *C. sonorensis* salivary gland homogenates (Perez de Leon & Tabachnick, 1996). In the *Culicoides* EST collection reported here, several putative 5′ nucleotidase apyrases were identified: two partial cDNAs with a metallophosphoesterase protein domain and one with a 5′ nucleotidase domain. Both of these domains may be found in a single full-length apyrase (Ribeiro *et al*., 2000b; Thomasova *et al*., 2002). All three *Culicoides* ESTs are most similar to *Anopheles gambiae* proteins.

A second putative platelet aggregation inhibitor was also identified. A conceptual translation of CssgEST00543 showed 60% amino acid (a.a.) identity and 71% similarity to *Drosophila melanogaster* platelet activating factor (PAF) acetylhydrolase alpha subunit (O9644), as well as 38% a.a. identity and 55% similarity to a rat isoform (NP_446106) (data not shown). PAF acetylhydrolase degrades PAF, a pro-inflammatory phospholipid that is activated during inflammatory injuries or infection (reviewed in Zimmerman *et al*., 2002). Although PAF hydrolysing activity has been found in salivary homogenates of *Culex quinquefasciatus* (Ribeiro & Francischetti, 2001), this is the first report of a PAF acetylhydrolase in the salivary gland transcriptome of a haematophagous disease vector.

New D7 family

The odorant/pheromone binding protein superfamily contains a class of proteins of unknown function referred to as the D7 family. This protein family was first identified as the most abundant class of secreted proteins expressed exclusively in salivary glands of female mosquitoes (James *et al*., 1991). The distinguishing feature among the generally dissimilar D7 and D7-related family members is a set of six conserved cysteines (Valenzuela *et al*., 2002a). These proteins have been proposed to function as small hydrophobic ligand carriers and/or in binding of host haemostatic factors (Valenzuela *et al*., 2002a). Table 2 shows a number of clusters that are proposed to encode new members of the D7 protein family. This group contains the most ESTs of the dataset, comprising 100 of the 708 ESTs analysed. With the exception of one, each of the corresponding proteins contains a signature protein domain that substantiates its place in the odorant/pheromone binding superfamily (Graham & Davies, 2002).

Table 3. Novel cDNAs constitute a new short D7 protein family

cDNA	MW	SP	MWm	pI
Cssg00004	15.3	16	13.6	4.6
Cssg00015	17.5	22	15.2	9.8
Cssg000211	18.5	21	16.2	9.6
Cssg00331	18.2	21	15.7	9.8
Cssg0642A	17.8	25	15.1	5.2
Cssg0642B	17.8	25	15.1	5.2
Cssg00666A	16.9	25	14.2	8.5
Cssg00666B	17.2	25	14.5	8.5
Cssg00673A	16.7	18	14.7	4.7
Cssg00673B	16.9	18	14.7	4.6

MW, predicted molecular weight; SP, putative signal peptide cleavage site (SignalP); MWm, predicted molecular weight of protein following signal peptide cleavage; pI, predicted isoelectric point.

Selected putative D7 proteins were aligned with *Ae. aegypti* D7Bclu1 protein in Fig. 2A. Three cysteines are conserved between mature *Ae. aegypti* D7Bclu1 and all mature *Culicoides* D7 proteins. Two additional cysteine residues are conserved among *Culicoides* D7 proteins but not with the *Aedes* protein, with a sixth cysteine in a variable position. Cssg00673A is the only one possessing eight cysteines. The predicted molecular weights of the mature *Culicoides* D7 proteins range from 13.6 to 16.2 kDa, 1.5–4.1 kDa smaller than the smallest previously reported D7 proteins (Table 3) (Valenzuela *et al*., 2002a). The predicted isoelectric points of the short D7 proteins are similar to short D7 in a variety of mosquito species, typically ranging from 4.6 to 5.08 or 8.2 to 9.66 (Valenzuela *et al*., 2002a). When considered with the female-specific expression patterns presented below, these shared characteristics support the inclusion of the *Culicoides* short D7 proteins as new members of the D7 class.

Salivary gland coagulation inhibitors

A wide variety of arthropods, including plant pests, express protease inhibitors for protection against pathogens or facilitation of feeding. Most blood-sucking arthropods secrete protease inhibitors to prevent coagulation and maximize the volume of blood collected (reviewed in Valenzuela, 2002). These inhibitors have convergently evolved to act at specific points in the tissue factor pathway, typically at thrombin or Factor Xa (FXa). One example of convergent evolution is found in blood FXa inhibitors that have evolved

independently in ticks, mosquitoes and midges. This type of protease inhibitor activity has been identified in *Culicoides* midges (Perez de Leon *et al*., 1998), black flies (Jacobs *et al*., 1990), ticks (Francischetti *et al*., 2002a) and culicine mosquitoes (Stark & James, 1996, 1998).

The midge salivary gland EST collection revealed a variety of abundant clusters with similarity to coagulation inhibitors. Among these, two clusters in Table 2 contain significant similarity to Kunitz-type protease inhibitors. Several additional abundant clusters represent novel proteins with domains less similar to typical Kunitz-type features. Three clusters bearing strong Kunitz-type domains are similar to tissue factor pathway inhibitor proteins (TFPI). Therefore, one or more of these cDNAs may be responsible for the FXa inhibitor activity previously identified in *Culicoides* salivary gland homogenates (Perez de Leon *et al*., 1998). Interestingly, the predicted translation product of *Culicoides* TFPI2 is most similar, with 32% identity and 38% similarity, to the tick Kunitz-type FXa inhibitor ixolaris (Francischetti *et al*., 2002a). The FXa inhibitor found in culicine mosquitoes is a highly diverged member of the serpin superfamily (Stark & James, 1998), and the black fly sequence has not been published. In contrast, anopheline mosquitoes secrete anticoagulants that act on thrombin (Stark & James, 1996).

A protein alignment showing the three *Culicoides* TFPI conceptual translations and ixolaris is shown in Fig. 2B. Localization of TFPI1 and TFPI2 showed selective accumulation of transcripts in adult female accessory salivary glands (Fig. 3); mRNA was not detected in any other adult female tissue (data not shown). Similar experiments using sense strand probe produced no signal (data not shown). This provides confirmation of tissue-specific transcript expression in the adult female midge. Supporting data for tissue-specific expression of TFPI3 is reported below.

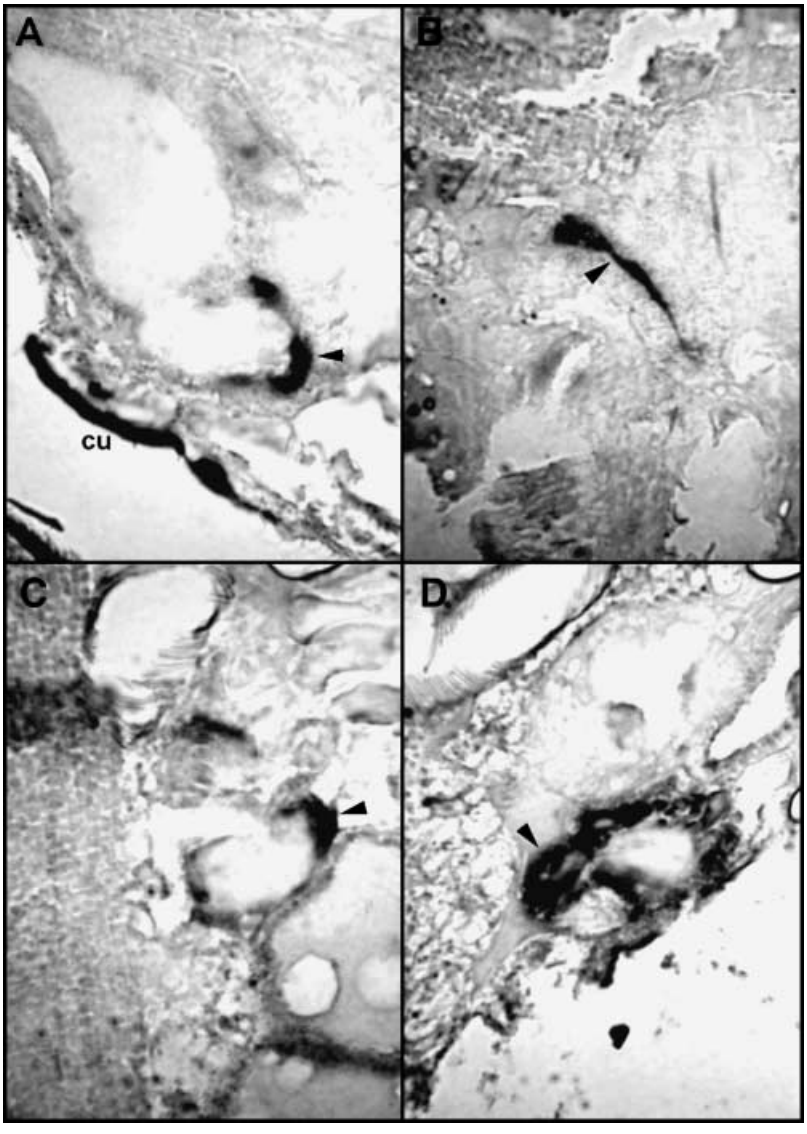


Figure 3. TFPI1 and TFPI2 transcripts localized in adult female blood-fed midge salivary accessory glands. Blood-fed female midges were paraffin-embedded, and digoxigenin-labelled probes were used for *in situ* localization of transcripts as indicated by arrows. Magnification 630×. (A) TFPI1 in female 2 days postfeeding. cu, cuticle. (B) TFPI1 in gravid female 3 days postfeeding. (C,D) TFPI2 localization in females 1 day postfeeding.

Salivary secreted enzymes

Three clusters of salivary factors responsible for sugar meal digestion were identified: amylase, maltase and aldolase (Table 2). In addition to these, sixteen ESTs code for two clusters of secreted serine proteases (Table 2). The distribution of one of these cDNAs, LTRYP3A, among different midge tissues as determined by quantitative PCR is described below.

Putative mucins

Recent reports of other disease vector salivary transcriptomes have taken note of the possible significance of secreted salivary mucins in the defence response (Francischetti *et al*., 2002b; Ribeiro *et al*., 2004). Therefore, all abundant *Culicoides* salivary EST clusters were checked for putative O-glycosylation sites using the NetOGlyc server (Hansen *et al*., 1998). Five consensus cDNAs with complete ORFs and poly adenylation signals code for predicted translation products with regions of low complexity (Table 2). All carry both a signal peptide and clustered serines or threonines indicative of putative mucins (Hansen *et al*., 1998). Four code for short proteins: Cssg00019, 63 a.a.; Cssg00656A, 81 a.a.; Cssg00656B, 80 a.a.; and Cssg00656C, 77 a.a. Cssg00021 is the longest with 168 a.a.

Quantitative real-time PCR of selected cDNAs

To corroborate the putative functional assignments of predicted proteins presented in Tables 1 and 2, quantitative real-time PCR (q-PCR) was performed for selected transcripts. Target tissues include heads with salivary glands, midguts, and remaining carcasses of both unfed and serum-fed adult female midges, as well as female salivary glands and whole males and females. This sample set was

Table 4. D7 and coagulation inhibitor transcript levels are higher in adult female salivary glands than other tissues; digestive proteases are enriched in midguts

A	SG00004	cl	SG00015	cl	SG00211	cl	SG00642	cl
Salivary glands	98.39	±42.55	45.15	±8.20	51.43	±29.38	51.84	±12.85
Unfed female	0.43	±0.07	0.45	±0.03	0.65	±0.40	1.22	±0.09
Unfed male	ND	ND	BD	–	0.15	±0.005	BD	–
Unfed female								
Head	11.24	±2.41	34.31	±4.22	46.72	±6.19	26.75	±2.59
Carcass	0.91	±0.59	0.41	±0.12	0.86	±0.37	0.34	±0.06
Midgut	ND	–	ND	–	BD	–	BD	–
Serum-fed female								
Head	12.55	±5.50	17.48	±1.25	23.18	±1.57	6.91	±0.68
Carcass	0.16	±0.04	0.38	±0.05	0.59	±0.10	0.09	±0.01
Midgut	ND	–	BD	–	0.07	±0.01	0.01	±0.003

B	SG00129	cl	SG00654	cl	SG00660	cl	TFPI3	cl
Salivary glands	51.96	±29.56	106.56	±83.21	103.44	±66.37	39.77	±10.27
Unfed female	0.76	±0.08	0.85	±0.04	0.73	±0.05	0.74	±0.04
Unfed male	BD	–	ND	–	ND	–	0.01	±0.04
Unfed female								
Head	21.21	±2.86	72.82	±8.83	42.24	±7.19	21.68	±0.38
Carcass	0.43	±0.18	0.18	±0.08	0.56	±0.14	0.46	±0.06
Midgut	ND	–	ND	–	ND	–	ND	–
Serum-fed female								
Head	17.85	±2.31	33.84	±3.91	20.72	±2.10	7.45	±0.02
Carcass	0.08	±0.02	0.20	±0.07	0.57	±0.08	0.21	±0.02
Midgut	ND	–	0.06	±0.01	BD	–	ND	–

C	TRY1_115	cl	TRY1_156	cl	CHYM1	cl	MDP6C	cl	LTRYP3A	cl	FER1	cl
Salivary glands	28.61	±19.06	ND	–	ND	–	ND	–	371.48	±41.05	1.19	±0.15
Unfed female	0.05	±0.02	2.83	±0.58	5.92	±0.27	BD	–	0.83	±0.12	1.74	±0.10
Unfed male	ND	–	BD	–	0.47	±0.06	BD	–	37.28	±2.53	4.51	±0.33
Unfed Female												
Head	0.05	±0.03	0.22	±0.26	0.43	±0.09	0.04	±0.05	19.43	±1.62	4.32	±0.78
Carcass	ND	–	10.19	±2.69	0.7	±0.13	ND	–	1.29	±0.60	11.59	±1.99
Midgut	0.24	±0.07	36.57	±9.11	51.10	±3.85	0.02	±0.001	0.02	±0.01	0.96	±0.09
Serum-fed female												
Head	ND	–	8.36	±10.60	2.19	±0.32	0.1	±0.02	26.16	±9.42	1.60	±0.19
Carcass	0.09	±0.03	0.64	±0.16	1.47	±0.12	0.12	±0.02	0.06	±0.02	2.69	±0.21
Midgut	3.01	±0.24	53.05	±5.86	51.29	±5.37	3.73	±0.34	0.03	±0.02	0.81	±0.43

Matched sets of adult female midge tissues before and within 5 h after a serum meal, as well as whole females and males and salivary glands were analysed by quantitative real-time PCR. This sample set was used for all transcripts analysed. Relative transcript levels are depicted as the mean of duplicate samples normalized to the level of ACT1D transcripts. 'ND', not detected; 'BD', below detection limits; 'cl', confidence interval. (A) D7 family transcript levels. (B) Putative protease inhibitor transcript levels. (C) Gut and salivary proteases; ferritin transcript levels.

assessed using an unfed adult female sample as a standard curve source for all q-PCR cDNA targets presented in Table 4, and thus allows comparison of relative transcript levels between transcripts, as well as between various tissues (except for MDP6C and TRY1_115: see Experimental procedures).

The hallmark of D7 protein family members is female salivary gland-specific protein expression (James *et al*., 1991; Arca *et al*., 1999). Here, we chose to assess transcript abundance of selected *Culicoides* D7 family cDNAs to provide corroborative evidence of their status in the D7 family. Relative transcript levels of Cssg00004, Cssg00015, Cssg00211 and Cssg00642 are presented in Table 4A. All candidates tested were most highly enriched in female salivary glands. Considered together, D7 family transcript levels were near or below detection limits in midguts, at low to moderate levels in all female carcasses (SD 0.46, range 0.04–1.53) and high levels in heads (SD 17.04, range 5.04–66.64). The presence of D7 transcripts in the carcass is probably due to the retention of residual salivary gland tissue remaining in the thorax upon head removal. With the exception of Cssg00004, all types of *Culicoides* D7 transcripts in Table 4A are depleted following feeding ($P < 0.05$, *t*-test).

When comparing relative D7 transcript levels to each other (Table 4A), Cssg00004 seemed to be the most abundant in purified salivary glands; this cDNA was also the single most abundant EST cluster in the salivary gland collection. Cssg00004, Cssg00015 and Cssg00642 levels were below detection limits in males (see Experimental procedures). Cssg00211 was present in low levels in males. Cssg00211 and Cssg00642 were present in low levels in serum-fed female midguts, suggesting possible roles in meal catabolism or defence. An additional D7 cDNA, Cssg00666A, was determined to be unsuitable for quantitative analysis due to the presence of multiple amplification fragments; however, the absence of detection in males supports its inclusion in the D7 family (data not shown).

Table 4B depicts the relative abundance of selected putative blood coagulation inhibitor cDNAs, either containing Kunitz-type protease inhibitor domains or weak homology to a venom-type trypsin inhibitor domain (Table 2). Considered together, all transcripts tested in Table 4B were depleted from female heads following a serum meal ($P < 0.05$, *t*-test). Cssg00129, Cssg00654, Cssg00660 and CsTFPI3 were highly enriched in adult female salivary glands and absent or below detection limits in males. Similar to the D7 protein family, one of the four cDNAs was present in low levels in adult female midguts following a serum meal, and all four were present in low levels in the carcass. As for other salivary factors, this may reflect residual gland retained in the carcass.

Transcript levels of the most abundant midgut serine protease EST clusters were assessed: TRY1_115, TRY1_156 and CHYM1. Together, these consensus transcripts represent about one-third of the serine protease clusters identified. The q-PCR analyses demonstrated that trypsin TRY1_115 is enriched in salivary glands (Table 4C) and increased in midguts significantly within a few hours of a serum meal ($P \leq 0.01$, *t*-test). In contrast, there were no significant differences in midgut levels of trypsin TRY1_156 and chymotrypsin CHYM1 transcripts following a serum meal, and both transcripts were absent from salivary glands. Whether meal-induced or constitutively expressed, all three of these transcripts were enriched in midguts over the remainder of the insect in the matched tissue sets, substantiating their roles as digestive proteases.

A serine protease identified in adult female salivary glands, LTRYP3A, presented a markedly different tissue distribution in that it was least abundant in midguts and most abundant in purified female salivary glands and adult males (Table 4C), suggesting a possible role in defence or plant feeding behaviour. Salivary gland secreted serine proteases have been described in both *Ae. aeygpti* and *Ixodes scapularis* (Valenzuela *et al*., 2002b,c). Further analysis will be required to determine whether *Culicoides* salivary serine proteases serve a role in post-translational modification of other proteins, non-haematophagous feeding or the defence response.

Two cDNA clusters were assessed for possible roles in meal catabolism, a metal-dependent protease (MDP6C) and a putative ferritin (FER1), an iron storage molecule. The predicted translation product of MDP6C has an astacin domain (1.9E-75) (Pfam website; Bateman *et al*., 2004) and shows 57% similarity and 43% identity to GmZmp, a midgut-specific metalloprotease of tsetse flies (Yan *et al*., 2002) (data not shown). When q-PCR Ct values were compared, MDP6C transcript levels are significantly different between serum-fed midguts and unfed midguts ($P \leq 0.01$, *t*-test), supporting its purported role in meal digestion.

The predicted translation product of *Culicoides* FER1 cDNA is most similar to *Ae. aegypti* ferritin light chain (LCH) with 49% a.a. similarity and 35% identity. As with other insect ferritins, *Culicoides* FER1 has a signal peptide (SignalP, Nielsen *et al*., 1997). Like the *Aedes* LCH sequence (Geiser *et al*., 2003), the *Culicoides* FER1 5′-untranslated region does not contain an iron responsive element (IRE) found in other ferritins (reviewed in Nichol *et al*., 2002); however, unlike *Aedes* LCH, *Culicoides* FER1 does not contain an N-glycosylation site. Table 4C shows that the FER1 transcript was present in all female midge tissues tested, as well as in males. Interestingly, transcript levels did not differ significantly between unfed and serum-fed midguts, as would be expected if FER1 was regulated at the translational level as occurs with other insect ferritins that contain an IRE (reviewed in Nichol *et al*., 2002). Considered together, the lack of an IRE in CsFER1 and the lack of FER1 transcript induction in the midge midgut following

a heme-deficient serum meal, along with the evidence that *Ae. aegypti* ferritin light and heavy chains are both regulated at the transcriptional level (Geiser *et al*., 2003), suggests that haematophagous insect midgut ferritin genes lacking an IRE are regulated at the transcriptional level by the presence of iron in the bloodmeal rather than at the translational level, as occurs with most non-haematophagous insect ferritins.

Interestingly, FER1 levels are higher in the unfed male than in the unfed female, and in female tissues, comparison of Ct values showed that levels of FER1 in the unfed female carcass are reduced within 5 h of feeding ($P \leq 0.01$, *t*-test). Therefore, *Culicoides* FER1 might also play a role in general iron homeostasis rather than merely protecting the midgut from the toxic effects of heme during bloodmeal catabolism. If so, additional regulatory elements probably control FER1 transcription in tissues other than the midgut. However, further characterization will be required fully to understand iron homeostasis in this insect.

Summary

This is the first report of substantial cDNA sequence information for the important disease vector *C. sonorensis*. The midgut cDNA library allowed identification of important conserved metabolic transcripts. The salivary gland cDNA library revealed a number of important antihaemostatic and putative immunomodulatory factors that may be used in future characterization of salivary potentiation of pathogen transmission by this insect. Two new classes of proteins were identified: a new class of D7 proteins and a new class of Kunitz-type protease inhibitors. The addition of these proteins to the growing dataset of dipteran salivary factors will help define the convergent evolution of blood-feeding insects.

Prior to this work, only fourteen nuclear-encoded mRNAs from *Culicoides* were in the public database, and eight of those were previously reported by this group (Campbell & Wilson, 2002). The cDNAs described here provide tools for future vector biology studies of biting midges and should facilitate future studies of arbovirus infection and replication. This collection may be used to elucidate possible genetic determinants of arbovirus infection and to devise potential strategies to control vector infection. In addition, this information should prove invaluable in developing novel control strategies for midge-transmitted diseases.

Experimental procedures

Insects

Culicoides sonorensis were reared at the colony maintained at the US Department of Agriculture Arthropod-Borne Animal Diseases Research Laboratory (ABADRL), Laramie, Wyoming. Adult females (AK colony), 2–3 days old, provided de-ionized water, were allowed to feed on a meal of fetal bovine serum, 100 mg/ml phenol red sodium salt (Sigma, St Louis, MO, USA), and phagostimulants ATP sodium salt, 50 mM, pH 7.0, and sodium bicarbonate, 0.37 mg/ml (Nunamaker *et al*., 2000). Midguts were removed into RNALater (Ambion, Austin, TX, USA) within 5 h postfeeding and stored at −80 °C. Salivary glands were removed from 2- to 5-day old females that had been provided sugar water since emergence; glands were stored in RNAlater at −20 °C.

cDNA libraries

Serum-fed midges were used to prevent isolation of blood cell RNA. Total RNA was extracted from approximately 218 midguts using the Micro Poly (A) Pure isolation (Ambion) and purified to generate poly A^+ RNA according to the manufacturer's recommendations. RNA was inserted into the plasmid pSPORT1, using the Gateway Technology plasmid system according to the manufacturer's recommendations (Invitrogen, Carlsbad, CA, USA).

For the salivary gland library, total RNA was extracted from approximately 300 glands using the RNAqueous-4-PCR (Ambion) extraction with DNase I treatment. Reverse transcription and cloning into the pDNR-LIB plasmid was performed using the CREATOR SMART system according to the manufacturer's recommendations (BD Biosciences, Palo Alto, CA, USA).

Single pass 5′ DNA sequencing was performed with T7 sequencing primers (GTAATACGACTCACTATAGGG) using the methods of Smith *et al*. (2000) on an ABI PRISM 3100 DNA analyser (Applied Biosystems, Foster City, CA, USA). Sequences were trimmed of vector and low-quality sequence using VNTI (Informax, Frederick, MD, USA) or Seqman software (DNASTAR, Madison, WI, USA). Consensus transcripts were determined using Seqman (DNASTAR).

Data analysis

In general, EST clusters were assigned putative function according to similarity to the most similar previously named cDNA or by the first letter of each syllable or word of the gene ontology molecular function term assigned to the most closely related cDNA by tblastx or blastx search. ORFs were identified using VNTI software (Informax) and confirmed by alignment to the most similar known cDNA by BLAST search (Altschul *et al*., 1990). Putative signal peptides were determined using SignalP (Nielsen *et al*., 1997), cutoff 0.7. Protein domains were identified using the Pfam website (Bateman *et al*., 2004) or the Conserved Domain Database (CDD; Marchler-Bauer *et al*., 2003). Putative O-glycosylation sites were determined using the NetOGlyc website (Hansen *et al*., 1998).

In situ *hybridization*

Colony midges (2–3 days old) were fed a blood meal with an artificial feeder according to established procedures (Hunt, 1994). Midges were held for 1–3 days, fixed and embedded in paraffin. cDNA fragments were amplified from a plasmid clone containing the insert of interest. PCR primers: TFPI1 forward CCGGGGATGATAACTTTT, TFPI1 reverse CCTCATTTCCACTTCCTT. The resulting cDNA insert was gel-purified, and *in vitro* transcription was performed to produce an antisense TFPI1 probe using a primer with 5′ extension bearing a T7 promoter (underlined): TAATACGACTCACTATAGGGAGACAAGGAAGTGGAAATG. A TFPI1 sense primer generated a negative control probe from the

same cDNA insert using a primer with a T3 promoter (underlined): AATTAACCCTCACTAAAGGGCTTGGGTCAAATCTCAATG. TFPI2 probes were prepared using similar methods. PCR primers: TFPI2 forward CGAGAGTCGTGCATTG, TFPI2 reverse TCCTCACAGCGACGAT. Antisense probe primer: TAATACGACTCACTATAGGGAATATTGCACGACCAACACC. Sense probe primer: AATTAACCCTCACTAAAGGGCCTTCAATTGTCAGGAGGA.

Digoxigenin probe preparation and *in situ* hybridization were performed according to the manufacturer's instructions (Roche Applied Science, Indianapolis, IN, USA) with the following modifications: prehybridization, 1 h, 42 °C; hybridization overnight, 42 °C. Probe was detected using NBT-BCIP substrate (Vector Laboratories, Burlingame, CA, USA) after 4 h of incubation. Sections were otherwise unstained.

Real-time quantitative PCR

RNA extractions. Adult female midges were fed a serum meal as described above. Pools of ten heads, midguts and remaining carcasses were dissected in RNAlater and transferred to tubes containing RNAlater. Similar dissections were performed for females that had only been provided water since eclosion. Whole insects were also processed in pools of ten. RNA extractions for all samples were prepared using RNAqueous-4-PCR (Ambion) extraction with DNase I treatment, according to the manufacturer's instructions. PCR was performed in the absence of reverse transcription to confirm the absence of detectable levels of contaminating genomic DNA. Reverse transcription reactions (RT rxns) were performed as described previously using 100 ng total RNA for each sample (Campbell & Wilson, 2002). RT rxns were diluted 1–4 for use in q-PCR.

q-PCR. Relative gene expression levels were determined for all cDNA targets using SYBR Green 2X Master Mix (Applied Biosystems, Foster City, CA) on an ABI PRISM 7000 (Applied Biosystems) according to the manufacturer's instructions. All samples were performed in quadruplicate. Dissociation curves were analysed for all cDNA targets to confirm a single peak, as multiple amplicons negate the accuracy of quantitative analysis. Cycling parameters: one cycle: 50 °C, 2 min, 95 °C, 10 min; forty cycles: 95 °C 15 s, 60 °C 1 min. A ten-fold serial dilution series was prepared from an RT rxn of unfed female midges to generate a standard curve for each sample run. This sample set was used for all cDNA targets analysed except for MDP6 and TRY1_115. For these two cDNAs, a similar standard dilution series was prepared from serum-fed midguts.

Relative gene expression analysis. ABI PRISM 7000 analytical settings were a 0.2 threshold value and automatic baseline calling for most targets in this study. Threshold settings and baseline settings were maintained across all independent runs for a given target cDNA. The standard curve method was used to determine relative transcript levels according to standard methods outlined in User Bulletin #2 (Applied Biosystems). Briefly, for each cDNA target, the standard curve slope and *y*-intercept were used to convert sample Ct values to estimated amounts of RNA. These values were normalized by a similar output value for actin generated from the same samples.

Statistical analysis. Reported values are the mean of duplicate samples. When an amplification signal was undetected, a value of 'zero' was assigned. When no amplification signal was detected in all four replicates, a value of 'ND' was assigned. Because SYBR Green detection of double-stranded amplicons is prone to false positive signals, rare targets that produced normalized values of ≤ 0.01 were designated as below detection levels (BD). Confidence intervals were calculated from the standard deviation of the normalized means using an alpha value of 0.05. Except where indicated, all *t*-tests were performed on normalized relative gene expression values. In some cases, proof of statistical significance required the comparison of raw Ct values to increase the apparent number of samples analysed, as normalized values represent compressed sample sets of multiple replicates.

Primers for q-PCR

These are as follows: CHYM1f55, TATATTCGCCATGTTGGCTTTG; CHYM1r113, TCCTTGGACTGATTTGTTTACC; Ltryp3f192, TGCCCGCTTGCATCAATTTC; Ltryp3r247, ACAACTTCCTCCTTCGCATGAc; MDP6Cf387, CATCAACTGTCGGAAGGAAAG; MDP6Cr470, TGGCAAGTCTGAAACAACc; SG0004f223, ATGGAAAGTGCCATCCAAc; SG0004r306, GATCCACCTGAACCAGATAAC; SG00015f249, ATGGATGCCCAAATTGGTAAC; SG00015r324, AAAGTGTGCGTAAATGAGg; SG00129f256, CAATGCGTCTATAAGTGTTGG; SG00129r384, CAGATCAGGAATTGGGATAGG; SG00211f371, GTCCAACACGAGACAACTG; SG00211r465, CCGCCATTGAAGTGAATCC; SG00642Af377, CATGAAAGTCGATGCCGATAG; SG00642Ar456, TGCCCAATTCAGCAGTTAC; SG00654Af122, GAGCCACGCCATAATCAAG; SG00654Ar211, GTCAGTTACCGGAACACAG; SG00660f286, TGCTGGGATTACAGGAAATG; SG00660r407, TTATCAGAGCGACTTGGAAC; SG00666Af344, GCCCAAGAAACACCTTCG; SG00666Ar407, CCGCTTCGTCAGTCAATC; CSTFPI3f147, TCCGAAGGATCGCAAATAAATG; CSTFPI3r249, CCTCCGACAACTTTCTTGTAAC; CSTRY1_115f98, GACAATTCAACGAGGATCTTTC; CSTRY1_115R152, TACGATACGTTCACCAAGAG; CSTRY1_156F763, GTTGTGCAGAGAAGAATTTCC; CSTRY1_156R825, TGATCCAATCACGAACGTAAG; CsFGS1F649, CTGGTGTTGCTCACTTTG; CsFGS1R763, CATACGGCTTGTCCTTTC; CSACT1DF686, TATGCCTTACCACATGCTATC; CCSACT1DF805, AATTTCACGTTCGGCAGTTG.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

Acknowledgements

This work is dedicated to the memory of Theodor Hanekamp, a researcher whose enthusiasm for science has outlived him in the hearts and minds of his students and colleagues. We thank J. Kempert and F. Stanek for providing colony-reared midges, M. Larson for dissecting midge midguts, and M. McNulty for dissecting salivary glands. We also thank M. West for helpful advice regarding statistical analysis. This work was performed with funding from the US Department of Agriculture, Agricultural Research service CRIS Project Number: 5410-32000-010-00D.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins,

R.A., Galle R.F., Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M., Mueller, H.M., Dimopoulos, G., Law, J.H., Wells, M.A., Birney, E., Charlab, R., Halpern, A.L., Kokoza, E., Kraft, C.L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G.M., Salzberg, S.L., Sutton, G.G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F.H., Ribeiro, J., Gelbart, W.M., Kafatos, F.C. and Bork, P. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Arca, B., Lombardo, F., Capurro, M., Della Torre, A., Spanos, L., Dimopoulos, G., Louis, C., James, A.A. and Coluzzi, M. (1999) Salivary gland-specific gene expression in the malaria vector *Anopheles gambiae*. *Parassitologia* **41**: 483–487.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.

Baker, K.P. and Quinn, P.J. (1978) A report on clinical aspects and histopathology of sweet itch. *Equine Vet J* **10**: 243–248.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C. and Eddy, S.R. (2004) The Pfam protein families database. *Nucleic Acids Res* **32** Database issue: D138–141.

Berridge, M.J., Buchan, P.B. and Heslop, J.P. (1984) Relationship of polyphosphoinositide metabolism to the hormonal activation of the insect salivary gland by 5-hydroxytryptamine. *Mol Cell Endocrinol* **36**: 37–42.

Budds, M., Edwards, J., Olaveson, A. and Gacesa, P. (1987) A comparison of the properties of the hyaluronidase from a temperate and a tropical species of leech. *Comparative Biochem Physiol* **87B**: 497–500.

Calvo, E., Andersen, J., De Francischetti, I.M.L.C.M., Debianchi, A.G., James, A.A., Ribeiro, J.M. and Marinotti, O. (2004) The transcriptome of adult female *Anopheles darlingi* salivary glands. *Insect Mol Biol* **13**: 73–88.

Campbell, C.L. and Wilson, W.C. (2002) Differentially expressed midgut transcripts in *Culicoides sonorensis* (Diptera: ceratopogonidae) following Orbivirus (reoviridae) oral feeding. *Insect Mol Biol* **11**: 595–604.

Cerna, P., Mikes, L. and Volf, P. (2002) Salivary gland hyaluronidase in various species of phlebotomine sand flies (Diptera: psychodidae). *Insect Biochem Mol Biol* **32**: 1691–1697.

Champagne, D.E., Smartt, C.T., Ribeiro, J.M. and James, A.A. (1995) The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5′-nucleotidase family. *Proc Natl Acad Sci USA* **92**: 694–698.

Charlab, R., Valenzuela, J.G., Rowton, E.D. and Ribeiro, J.M. (1999) Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc Natl Acad Sci USA* **96**: 15155–15160.

Francischetti, I.M., Valenzuela, J.G., Andersen, J.F., Mather, T.N. and Ribeiro, J.M. (2002a) Ixolaris, a novel recombinant tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick, *Ixodes scapularis*: identification of factor X and factor Xa as scaffolds for the inhibition of factor VIIa/tissue factor complex. *Blood* **99**: 3602–3612.

Francischetti, I.M., Valenzuela, J.G., Pham, V.M., Garfield, M.K. and Ribeiro, J.M. (2002b) Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector Anopheles gambiae. *J Exp Biol* **205**: 2429–2451.

Geiser, D.L., Chavez, C.A., Flores-Munguia, R., Winzerling, J.J. and Pham, D.Q. (2003) *Aedes aegypti* ferritin. *Eur J Biochem* **270**: 3667–3674.

Gillespie, R.D., Dolan, M.C., Piesman, J. and Titus, R.G. (2001) Identification of an IL-2 binding protein in the saliva of the Lyme disease vector tick, *Ixodes scapularis*. *J Immunol* **166**: 4319–4326.

Graham, L.A. and Davies, P.L. (2002) The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene* **292**: 43–55.

Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* **15**: 115–130.

Henriksen, A., King, T.P., Mirza, O., Monsalve, R.I., Meno, K., Ipsen, H., Larsen, J.N., Gajhede, M. and Spangfort, M.D. (2001) Major venom allergen of yellow jackets, Ves v, 5: structural characterization of a pathogenesis-related protein superfamily. *Proteins* **45**: 438–448.

Holt, R., Subramanian, A., Halpern, G.M., Sutton, A., Charlab, G.G., Nusskern, R., Wincker, D.R., Clark, P., Ribeiro, A.G., Wides, J.M., *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Hunt, G. (1994) *A Procedural Manual for the Large-Scale Rearing of the Biting Midge, Culicoides variipennis (Diptera: Ceratopogonidae).* US Department of Agriculture, Agricultural Research Service, Laramie, Wyoming, USA.

Jacobs, J.W., Cupp, E.W., Sardana, M. and Friedman, P.A. (1990) Isolation and characterization of a coagulation factor Xa inhibitor from black fly salivary glands. *Thromb Haemost* **64**: 235–238.

James, A.A., Blackmer, K., Marinotti, O., Ghosn, C.R. and Racioppi, J.V. (1991) Isolation and characterization of the gene expressing the major salivary gland protein of the female mosquito, *Aedes aegypti*. *Mol Biochem Parasitol* **44**: 245–253.

Jones, L.D., Matthewson, M. and Nuttall, P.A. (1992) Saliva-activated transmission (SAT) of Thogoto virus: dynamics of SAT factor activity in the salivary glands of *Rhipicephalus appendiculatus*, *Amblyomma variegatum*, and *Boophilus microplus* ticks. *Exp Appl Acarol* **13**: 241–248.

Lehane, M.J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J., Soares, M.B., Bonaldo, M.F., Lehane, S. and Hall, N. (2003) Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. *Genome Biol* **4**: R63.

Li, S., Kwon, J. and Aksoy, S. (2001) Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*. *Insect Mol Biol* **10**: 69–76.

Limesand, K.H., Higgs, S., Pearson, L.D. and Beaty, B.J. (2000) Potentiation of vesicular stomatitis New Jersey virus infection in mice by mosquito saliva. *Parasite Immunol* **22**: 461–467.

Limesand, K.H., Higgs, S., Pearson, L.D. and Beaty, B.J. (2003) Effect of mosquito salivary gland treatment on vesicular stomatitis New Jersey virus replication and interferon alpha/beta expression in vitro. *J Med Entomol* **40**: 199–205.

Lu, G., Kochoumian, L. and King, T.P. (1995) Sequence identity and antigenic cross-reactivity of white face hornet venom

allergen, also a hyaluronidase, with other proteins. *J Biol Chem* **270**: 4457–4465.

Marchler-Bauer, A., Anderson, J.B., Deweese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., *et al.* (2003) CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* **31**: 383–387.

Mckee, C.M., Penno, M.B., Cowman, M., Burdick, M.D., Strieter, R.M., Bao, C. and Noble, P.W. (1996) Hyaluronan (HA) fragments induce chemokine gene expression in alveolar macrophages. The role of HA size and CD44. *J Clin Invest* **98**: 2403–2413.

Mckee, C.M., Lowenstein, C.J., Horton, M.R., Wu, J., Bao, C., Chin, B.Y., Choi, A.M. and Nobie, P.W. (1997) Hyaluronan fragments induce nitric-oxide synthase in murine macrophages through a nuclear factor kappaB-dependent mechanism. *J Biol Chem* **272**: 8013–8018.

Mckeever, S., Brickle, D.S. and Hagan, D.V. (1997) Mouthparts, antennae and genitalia of intersex *Culicoides stellifer* parasitized by mermithid nematodes. *Med Vet Entomol* **11**: 217–222.

Mellor, P.S., Boorman, J. and Baylis, M. (2000) Culicoides biting midges: their role as arbovirus vectors. *Annu Rev Entomol* **45**: 307–340.

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science* **287**: 2196–2204.

Nakazawa, H., Tsuneishi, E., Ponnuvel, K.M., Furukawa, S., Asaoka, A., Tanaka, H., Ishibashi, J. and Yamakawa, M. (2004) Antiviral activity of a serine protease from the digestive juice of *Bombyx mori* larvae against nucleopolyhedrovirus. *Virology* **321**: 154–162.

Nichol, H., Law, J.H. and Winzerling, J.J. (2002) Iron metabolism in insects. *Annu Rev Entomol* **47**: 535–559.

Nielsen, H., Engelbrecht, J., Brunak, S. and Von Heijne, G. (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**: 581–599.

Norsworthy, N.B., Sun, J., Elnaiem, D., Lanzaro, G. and Soong, L. (2004) Sand fly saliva enhances *Leishmania amazonensis* infection by modulating interleukin-10 production. *Infect Immun* **72**: 1240–1247.

Nunamaker, R.A., Perez De Leon, A.A., Campbell, C.L. and Lonning, S.M. (2000) Oral infection of *Culicoides sonorensis* (Diptera: Ceratopogonidae) by vesicular stomatitis virus. *J Med Entomol* **37**: 784–786.

Perez de Leon, A.A. and Tabachnick, W.J. (1996) Apyrase activity and adenosine diphosphate induced platelet aggregation inhibition by the salivary gland proteins of *Culicoides variipennis*, the North American vector of bluetongue viruses. *Vet Parasitol* **61**: 327–338.

Perez de Leon, A.A., Ribeiro, J.M., Tabachnick, W.J. and Valenzuela, J.G. (1997) Identification of a salivary vasodilator in the primary North American vector of bluetongue viruses, *Culicoides variipennis*. *Am J Trop Med Hyg* **57**: 375–381.

Perez de Leon, A.A., Valenzuela, J.G. and Tabachnick, W.J. (1998) Anticoagulant activity in salivary glands of the insect vector *Culicoides variipennis sonorensis* by an inhibitor of factor Xa. *Exp Parasitol* **88**: 121–130.

Quinn, P.J., Baker, K.P. and Morrow, A.N. (1983) Sweet itch: responses of clinically normal and affected horses to intradermal challenge with extracts of biting insects. *Equine Vet J* **15**: 266–272.

Ribeiro, J.M. (2003) A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal. *Insect Biochem Mol Biol* **33**: 865–882.

Ribeiro, J.M. and Francischetti, I.M. (2001) Platelet-activating-factor-hydrolyzing phospholipase C in the salivary glands and saliva of the mosquito *Culex quinquefasciatus*. *J Exp Biol* **204**: 3887–3894.

Ribeiro, J.M., Charlab, R., Rowton, E.D. and Cupp, E.W. (2000a) *Simulium vittatum* (Diptera: Simuliidae) and *Lutzomyia longipalpis* (Diptera: Psychodidae) salivary gland hyaluronidase activity. *J Med Entomol* **37**: 743–747.

Ribeiro, J.M., Rowton, E.D. and Charlab, R. (2000b) The salivary 5′-nucleotidase/phosphodiesterase of the hematophagus sand fly, *Lutzomyia longipalpis* [corrected]. *Insect Biochem Mol Biol* **30**: 279–285.

Ribeiro, J.M., Charlab, R., Pham, V.M., Garfield, M. and Valenzuela, J.G. (2004) An insight into the salivary transcriptome and proteome of the adult female mosquito *Culex pipiens quinquefasciatus*. *Insect Biochem Mol Biol* **34**: 543–563.

Smith, T.P., Godtel, R.A. and Lee, R.T. (2000) PCR-based setup for high-throughput cDNA library sequencing on the ABI 3700 automated DNA sequencer. *Biotechniques* **29**: 698–700.

Stark, K.R. and James, A.A. (1996) Salivary gland anticoagulants in culicine and anopheline mosquitoes (Diptera: Culicidae). *J Med Entomol* **33**: 645–650.

Stark, K.R. and James, A.A. (1998) Isolation and characterization of the gene encoding a novel factor Xa-directed anticoagulant from the yellow fever mosquito, *Aedes aegypti*. *J Biol Chem* **273**: 20802–20809.

Thomasova, D., Ton, L.Q., Copley, R.R., Zdobnov, E.M., Wang, X., Hong, Y.S., Sim, C., Bork, P., Kafatos, F.C. and Collins, F.H. (2002) Comparative genomic analysis in the region of a major Plasmodium-refractoriness locus of *Anopheles gambiae*. *Proc Natl Acad Sci USA* **99**: 8179–8184.

Titus, R.G. and Ribeiro, J.M. (1988) Salivary gland lysates from the sand fly *Lutzomyia longipalpis* enhance Leishmania infectivity. *Science* **239**: 1306–1308.

Valenzuela, J.G. (2002) High-throughput approaches to study salivary proteins and genes from vectors of disease. *Insect Biochem Mol Biol* **32**: 1199–1209.

Valenzuela, J.G., Charlab, R., Gonzalez, E.C., De Miranda-Santos, I.K., Marinotti, O., Francischetti, I.M. and Ribeiro, J.M. (2002a) The D7 family of salivary proteins in blood sucking diptera. *Insect Mol Biol* **11**: 149–155.

Valenzuela, J.G., Francischetti, I.M., Pham, V.M., Garfield, M.K., Mather, T.N. and Ribeiro, J.M. (2002b) Exploring the sialome of the tick *Ixodes scapularis*. *J Exp Biol* **205**: 2843–2864.

Valenzuela, J.G., Pham, V.M., Garfield, M.K., Francischetti, I.M. and Ribeiro, J.M. (2002c) Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem Mol Biol* **32**: 1101–1122.

Valenzuela, J.G., Francischetti, I.M., Pham, V.M., Garfield, M.K. and Ribeiro, J.M. (2003) Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. *Insect Biochem Mol Biol* **33**: 717–732.

Wang, W.H., Lu, J.X., Yao, P., Xie, Y. and Huang, Z.X. (2003) The distinct heme coordination environments and heme-binding stabilities of His39Ser and His39Cys mutants of cytochrome b5. *Protein Eng* **16**: 1047–1054.

Yan, J., Cheng, Q., Li, C.B. and Aksoy, S. (2002) Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. *Insect Mol Biol* **11**: 57–65.

Yeruham, I., Braverman, Y. and Orgad, U. (1993) Field observations in Israel on hypersensitivity in cattle, sheep and donkeys caused by *Culicoides*. *Aust Vet J* **70**: 348–352.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M., Mueller, H.M., Dimopoulos, G., Law, J.H., Wells, M.A., Birney, E., Charlab, R., Halpern, A.L., Kokoza, E., Kraft, C.L., Lai, Z., Lewis, S., Louis, C., Barillas-Mury, C., Nusskern, D., Rubin, G.M., Salzberg, S.L., Sutton, G.G., Topalis, P., Wides, R., Wincker, P., Yandell, M., Collins, F.H., Ribeiro, J., Gelbart, W.M., Kafatos, F.C. and Bork, P. (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**: 149–159.

Zimmermann, B. (2000) Control of InsP3-induced Ca^{2+} oscillations in permeabilized blowfly salivary gland cells: contribution of mitochondria. *J Physiol* **525**: 707–719.

Zimmerman, G.A., Mcintyre, T.M., Prescott, S.M. and Stafforini, D.M. (2002) The platelet-activating factor signaling system and its regulators in syndromes of inflammation and thrombosis. *Crit Care Med* **30**: S294–S301.